REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1999	Final Proceedings (1 Sep 98 - 31 Dec 98)

4. TITLE AND SUBTITLE
DNA Microarray Workshop

5. FUNDING NUMBERS
DAMD17-98-1-8660

6. AUTHOR(S)
Mr. Bernie Lauro

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Genomic Research
Rockville, Maryland 20850

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

DTIC QUALITY INSPECTED 4

19990219014

14. SUBJECT TERMS	15. NUMBER OF PAGES	16. PRICE CODE
	10	

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

AD_____

Award Number DAMD17-98-1-8660

TITLE: DNA Microarray Workshop

PRINCIPAL INVESTIGATOR: Mr. Bernie Lauro

CONTRACTING ORGANIZATION: Institute for Genomic Research
Rockville, Maryland 20850

REPORT DATE: January 1999

TYPE OF REPORT: Final Proceedings

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

Workshop on Functional Analysis of the Malaria Genome

The Institute for Genomic Research
November 9-10, 1998

On Monday, November 9 and Tuesday, November 10, 1998, The Institute for Genomic Research hosted along with The Department of Defense, a Workshop on Functional Analysis of the Malaria Genome.

The workshop was the first of three planned workshops that arose from discussions at the Malaria Genome Project meeting held in Cambridge, UK on 1-3 July 1998. Funding for the workshop was provided by the US Department of Defense (USAMRAA). The aim of the meeting was to discuss how to best use the data derived from the Malaria Genome Sequencing Project for the functional analysis of the genome. The workshop was chaired jointly be Robert Strausberg (NCI) and Daniel Carucci (NMRC). Thirty participants attended the workshop including representatives from the sequencing centers (Stanford University sent a statement in lieu of a representative), the funding agencies, malaria and other parasitic research labs and individuals with experience in emerging gene expression technologies.

Claire Fraser, President and Director of The Institute for Genomic Research opened the meeting and spoke of the importance of moving forward with expression analysis derived from genome sequence data. Representatives from each of the Malaria Genome Project Funding agencies also provided brief comments. In particular, Mike Gottlieb (NIAID) spoke of the new Malaria Repository at the American Type Tissue Collection (ATCC) and funding opportunities from this new NIH venture.

Dan Lawson (Sanger) reported that chromosome 3 was in 8 segments and 4 segments representing 480 kb of sequence were finished. The largest gap was approximately 6 kb. The approach to gap closure would include transposon-based techniques and reprobing PUC libraries. Annotation using Genefinder has revealed 35-40% positive BLAST hits on the chromosome 3 data. He also provided an update of the other chromosome projects and their status.

Malcolm Gardner (TIGR) reported on the completion of chromosome 2 and publication in Science. He also reported that the random phase of chromosome 14 was completed and that gap closure would be started shortly. Test libraries for chromosome 11 had been completed and were evaluated.

Stanford University was not present for an update of the chromosome 12 project. (Richard Hyman provided input to DNA microarrays)

Assessing gene and protein expression in Plasmodium: challenges and solutions.

Alister Craig (Oxford, UK) described the use of array technology to address questions based on the var genes of *P. falciparum*. These genes encode a variant antigen (PfEMP1) expressed on the surface of infected erythrocytes which has been implicated in the pathogenesis of malaria. Highly polymorphic regions of var genes from both laboratory and field isolates have been amplified using degenerate primers and will be arrayed onto nylon membranes for hybridisation analysis. These will be used to address questions on the diversity of repertoires between different strains of *P.falciparum* as well as looking at the expression patterns of var genes in infected populations.

Dan Carucci (NMRC) stressed the importance of the "next genome project" as one which utilizes sequence data for the development of new vaccines and drugs. He introduced the concept of genome-wide gene expression and reviewed several of the available methods for gene expression analyses, including DNA microarrays, SAGE, differential display, amplification-based strategies as well as characterising proteins. He reviewed some of the problems particularly associated with Plasmodium including poor access to abundant material from much of the like cycle and suggested that alternate strategies would be needed to examine gene and protein expression from much of the parasite life cycle.

DNA Microarray Technology

John Quackenbush (TIGR) discussed the relative advantages and disadvantages of various gene expression systems, particularly DNA microarrays and oligoarrays (Affymetrix). He reviewed TIGR microarray program, initially as part of the Molecular Dynamics Technology Access (TA) Program. TIGR has established a strong infrastructure including microarray robotics capability having purchased a state-of-the-art robotic microarrayer system from Intelligent automation, slide scanner system (Molecular Dynamics Generation III scanner and a General Scanning ScanArray 3000 scanner), relational database systems and software engineers both in house and through external collaborations. He demonstrated recent TIGR microarray data including that obtained using novel glass surfaces under development with an industrial partner. He demonstrated software developed in-house to display microarray data using a Java-based system. He reviewed several of the microarray projects his groups is working on, including a collaboration with NMRC on *P. falciparum* chromosome-specific DNA microarrays.

Maryanne Vahey (WRAIR) provided overview of their work with oligoarray technology. Using the Affymetrix oligochip technology in use for HIV research. The WRAIR Array laboratory has explored the use of high density oligo arrays to monitor the expression of genes associated yeast reporters of drug resistance to putative anti-malarial drugs. She reported that mining of the yeast malaria data sets will generate concrete hypotheses concerning the interaction of expressed genes both individually and as networks. These hypotheses could then be tested by perturbing the yeast strains by exposing them to antimalarial drugs and reassessing the gene expression patterns.

Swati Patankar (Harvard) discussed their groups efforts in gene expression analysis using the Serial Analysis of Gene Expression (SAGE) technology an alternative method of assessing changes in gene expression in *Plasmodium*. SAGE relies on enzymatic manipulations of cDNA resulting in short tags that can be used to identify genes and also to quantitate levels of gene expression. They are in the process of modifying the existing SAGE protocol that is optimized for human cDNA to make it more suitable for the A-T rich nature of *Plasmodium* cDNA.

Genomes to Vaccines

Stephen Hoffman (NMRC) discussed the use of genomic sequence data for the development of vaccines. He emphasized the need for two "extremes" of vaccines; one type for the non-immune travelers (liver stage vaccines) and one type for children in the developing world (blood stage vaccines). He outlined a pilot proposal to use data from the Malaria Genome Project for the development of a multistage malaria DNA vaccine. The proposal would utilize data from the completed *P. falciparum* chromosome 2 project to immunize mice with DNA vaccines designed against each gene, generate anti-sera to the proteins encoded by each gene in order to screen for protein expression in malaria parasites by immunofluorescent antibody testing.

WRAIR - Keith Martin outlined the strategy undertaken at WRAIR to utilize oligoarrays (see Vahey above) and DNA microarrays produced in conjunction with Dan Goldberg (WashU) and Mike Bittner (NHGRI) to study gene expression in *P. falciparum*. They plan to produce DNA microarrays estimated to cover 50% of the genes using combined GST/ESTs from clones HB3 and DD2 and from OFRs from clone 3D7 (Malaria Genome Project) available in public databases. He outlines several technical pitfalls including the co-precipitation of contaminating plasmid with certain mini prep methods as well as contaminating carbohydrates during RNA precipitation. He also reviewed the current status of the Affymetrix platform using yeast parental and knockout strains (CQ refractory and CQ susceptible) to validate the system.

NMRC/TIGR – Dan Carucci reviewed the chromosome-specific DNA microarray pilot project and results from a chromosome 2 DNA microarray. He discussed the advantages of a chromosome-specific array project which avoids the complications and pitfalls of using universal primers and plasmid templates for array construction as each gene is amplified with a specific pair of primers. He discussed the potential for misidentification

of genes through retrieval error which can occur in up to 30% of the clones (IMAGE project) or in 10% of recent experiences (see Richard Hyman below). He discussed the advantages of fully integrating a DNA microarray project into genome sequencing by using the power of relational databases (see John Quackenbush below). He presented results of a chromosome 2-specific DNA microarray hybridized to highly synchronized ring and schizont stage parasites and discussed the importance of data management for the project. He also showed an amplification-based strategy using either TaqMan or interchelating dyes for real time quantitative PCR.

Stanford – Richard Hyman was not in attendance, however Victoria McGovern provided an e-mail from Richard outlining his thoughts on arrays. He expressed concern regarding potential misidentification of genes using pre-existing DNA, either EST/GSTs or sequencing plasmids, as templates for array construction and suggested that differing PCR product lengths and base composition may confound interpretation of data. He felt that unique primers for each gene should be constructed taking into account product length and base composition. For oligoarrays, in situ synthesis ensures normalization of oligo length and composition.

Cath U/Stanford – Pradip Rathod reviewed the pilot work he is conducting with Pat Brown using a newly constructed GST (Mung Bean) library using *P. falciparum* clone 7G8 prepared by David Kaslow (NIH). The array contains 3648 spots from the library, although the level of redundancy was not determined. Hybridization with sexual stage and asexual stage RNA has identified 6 – 12 genes which are over-expressed in sexual stage parasites. These genes are being sequenced individually. Pradip demonstrated a web interface to the database at Stanford in real time on line to demonstrate the power of array databases

Utilization of expression data

Alan Cowman (WEHI) emphasized the importance of increasing the efficiency of transfection and developing an inducible promoter system as well as the identification of new selectable genes. Although currently transformation efficiency is low and it is difficult to disrupt genes yielding deleterious phenotypes, transfection is possible. The development of an inducible promoter may help in gene deletions that are lethal. He suggested several short term goals: (1) increase emphasis on improving transfection technology and efficiency, (2) conduct a pilot project with chromosome 2 knockouts of non-essential genes (genes not expressed in asexual stage parasites), and (3) doing knockouts of essential genes placed under control of an inducible promoter. He noted that the most important aspect of moving transfection into more labs is robust cultivation capability and not transfection technology.

David Baker (LSHTM) underscored the importance of understanding the steps in malaria parasite life cycle differentiation which at the molecular level are poorly understood. He outlined his work on cyclic nucleotides (cAMP/cGMP) thought to be important in controlling differentiation in malaria parasites; cAMP (adenylyl cyclase) is thought to trigger gametocytogenesis and cGMP (guanylyl cyclase) exflagellation. He provided a

description of the isolation of 2 cyclase genes from *P. falciparum* including: IFAT, EM localization, structural features etc and discussed results of transfections with knockout constructs incorporating these 2 genes and 2 other sexual stage genes in collaboration with Dr. Alan Cowman. He noted that completion of the Malaria Genome Sequencing project will allow identification of all the potential components of signaling pathways involved in differentiation and proposed to use microarray technology to identify the sexual stage gene repertoire. Gene knockout studies can then be used for functional analysis and verification of the predicted pathways.

John Kelly (LSHTM) discussed the functional analysis of the diploid *Trypanosoma cruzi* genome highlighting the availability of robust genetic manipulation techniques for this organism. He suggested that worker in this field might serve as a model for future work with *Plasmodium*. In *T. cruzi*, there are several different genetic tools in use, including episomal vectors which can facilitate overexpression of defined genes, or modified genes, and cosmid shuttle vectors which allow the introduction of large fragments of DNA into transfected parasites. He noted that integration occurs almost exclusively by homologous recombination. Their work includes the development of negative selection, and inducible expression system and chromosomal knockouts.

Database and software development

John Quackenbush (TIGR) reviewed their work of integration of microarray data with the sequencing relational databases (SyBase) in use at TIGR. He demonstrated laboratory software tools that aid in managing microarray projects and data visualization tools for data mining. Alex Saeed (TIGR) demonstrated "Array Viewer" a TIGR-developed Java-based cross platform application that can read flat files or database stored microarray information and display this data in a graphical format. Future improvements to Array Viewer are 1) expression ratio statistics, 2) alternate color schemes, 3) improved linear regression statistics, 4) alternate normalized schemes.

Other

Yemin Wu (ATCC) introduced the NIH sponsored Malaria Repository located at the ATCC as a means for researchers to have improved access to parasite, vector and human reagents, and to standardize assays using well characterized and renewable reagents. The goal is to relieve investigators of the burden of providing reagents to other investigators. An advisory committee will be established to interface with the malaria community. In addition there are plans for workshops on cultivation, transfection, micromanipulation and others.

Summary of recommendations - ***no formal votes were taken on any of the recommendations; rather there appeared to be consensus on several of the issues discussed by the participants.***

There seemed to be a consensus that microarray technology should be decentralized so that more individual laboratories would have access to the technology. Although printing technology should remain centralized due to the high purchase costs of arrayers, the more reasonable price of scanning hardware should allow moderately funded labs to perform and analyze microarray hybridization experiments in their labs. Labs will have to invest in training and personnel to maximize the benefits of having the technology in house. This

could be done by a centralized training facility (?Malaria Repository function?). Funding should be made available to microarray pilot projects.

The relative merits of the various gene expression technologies were discussed and in was generally felt that although arrays are being produced from GST/EST libraries from a variety of parasite clones, that gene specific arrays should be made as validated sequence data is available. With the completed genome, Affymetrix oligo arrays become more attractive though still more expensive than DNA microarrays. Other gene expression systems, such as SAGE should also be considered once some preliminary data is generated.

Most thought that there should be open access to various libraries, cDNA, GST, and large insert libraries (such as that from Peter DeJong's lab).

There was a strong feeling that improved molecular tools were needed, including transfection, knockouts and others to facilitate use of any data generated from gene expression technologies.

Although the techniques available to studying gene expression for blood stage parasites appear to be sufficiently robust to begin generating data soon (Cath U and NMRI/TIGR), there was a concern that not enough is being done to look at gene and protein expression from mosquito and liver stage parasites. Several ideas were presented, including an amplification-based strategies, but others discussed using methods to characterize protein expression, for example using DNA vaccines to generate antibodies as a means to detect protein expression. It was suggested that work done in the cancer field looking at protein expression in neoplastic cells might be applicable parasite infected liver cells.

There was a lengthy discussion regarding the ideal reagents for constructing arrays, including gene specific PCR products, arrayed sequencing clones, PCR products derived from EST/GST/sequencing libraries. No clear consensus was reached, but it was felt that results from the pilots should provide some clearer direction.

Several other issues were raised, such as the need for improved data mining tools, access to 96-well oligo synthesizers, and focus groups for vaccines and drugs.

Additional Information

For your additional information and reference, a list of attendees and agenda are sent with this report.

Attendee list

David Baker
Daniel Carucci, Co-chair
Alan Cowman
Alister Craig
Leda Cummings
Alexandra Fairfield
Michael Ferdig
Claire Fraser
Malcolm Gardner
Pat Goodwin
Michael Gottlieb
Rhian Hayward
Steve Hoffman
John Kelly
Dennis Kyle
Daniel Lawson
Keith Martin
Victoria McGovern
Lance Miller
Ian Morris
Peter Mylar
Swati Patankar
John Quackenbush
Pradip Rathod
Bill Rogers
Robert Strausberg, Co-chair
Zin-Zhuan Su
Herve Tettelin
Maryanne Vahey
Adam Witney

WORKSHOP ON FUNCTIONAL ANALYSIS OF THE MALARIA GENOME

THE INSTITUTE FOR GENOMIC RESEARCH, ROCKVILLE, MD
November 9-10, 1998

LOGISTICS FACTSHEET

The Department of Defense and the TIGR Conference Department looks forward to your participation at the Workshop on Functional Analysis of the Malaria Genome. The following information is provided to help you with your planning for the workshop. If you have any questions, please do not hesitate to contact Ms. Bernie Lauro at 301-838-3568.

The design of this workshop is purposely informal. Individuals are encouraged to present brief discussions of their field of research and to share early data and results where appropriate. Adequate time will be allowed for such presentations. Audiovisual facilities (overhead and 35mm slide projector as well as a computer LCD projector) will be provided.

AGENDA:

0900 Welcome and Introduction
0930 Summary of the Malaria Genome Sequencing Project
  Progress Report from Sanger
  Progress Report from Stanford
  Progress Report from TIGR
1130 Assessing gene and protein expression in *Plasmodium*: challenges and solutions
1230 Lunch provided
1400 DNA microarray technology
1530 Alternative methods of expression analysis
1700 Adjourn
1800 Reception at TIGR

10 November, 1998

0900 Review of microarray pilot projects
  Progress Report from Stanford
  Progress Report from TIGR
1030 Establishing directions and goals
1130 Database and software development
1230 Lunch provided
1330 Setting the agenda for future work
1700 Adjourn

MEALS: The Workshop will provide lunch on the two days of the conference and an evening reception on Monday evening, November 9, 1998.

TIGR Conference Department ■ 9712 Medical Center Drive ■ Rockville, MD 20850
301-838-3561 ■ FAX: 301-838-0229 ■ http://www.tigr.org/conf